PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :	A1	(11) International Publication Number: WO 00/05852
H04L 29/06		(43) International Publication Date: 3 February 2000 (03.02.00)

(21) International Application Number: PCT/US99/12394

(22) International Filing Date: 20 July 1999 (20.07.99)

(30) Priority Data:
60/093,554 21 July 1998 (21.07.98) US

(71) Applicant: RAYTHEON COMPANY [US/US]; 141 Spring Street, Lexington, MA 02421 (US).

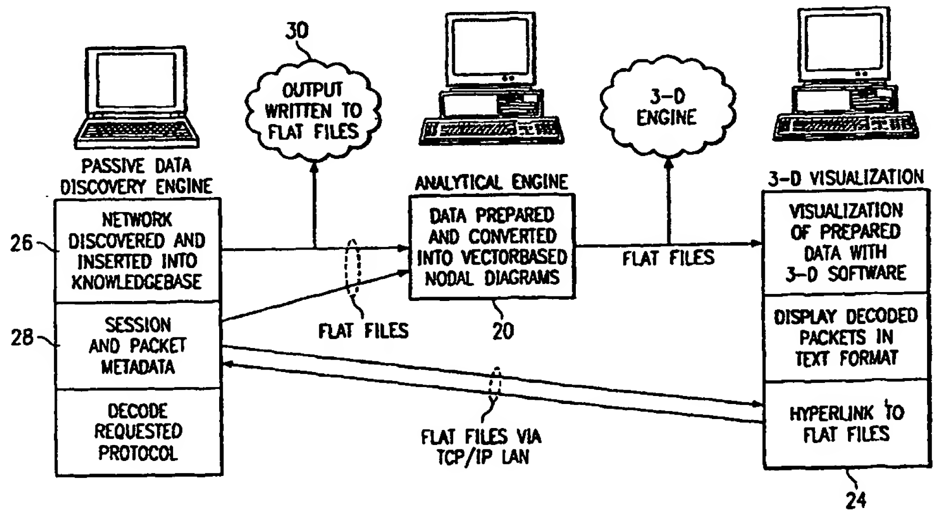
(72) Inventors: MALONEY, Michael, P.; 248 Poe Court, Severna Park, MD 21146 (US). SUIT, John, M.; 7873 Dero Drive, Pasadena, MD 21122 (US). SCOTT, Christopher, J.; 265 Steeplechase Drive, Pleasant Gap, PA 16823 (US). WOODUS, Francis, M.; 9733 Sherwood Farm Road, Owings Mill, MD 21117 (US). RUBEL, Rich; 205 Fox Ridge Court, Glen Burnie, MD 21061 (US). KAROLCHIK, Joseph; 890 The Old Station Court, Woodbine, MD 21797-8762 (US).

(74) Agents: MEIER, Harold, E. et al.; Baker & Botts, L.L.P., 2001 Ross Avenue, Dallas, TX 75201-2980 (US).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INFORMATION SECURITY ANALYSIS SYSTEM



(57) Abstract

The analysis system is a collection, configuration and integration of software programs that reside on multiple interconnected computer platforms. The software, less computer operating systems, is a combination of sensor, analysis, data conversion, and visualization programs. The hardware platforms consist of several different types of interconnected computers, which share the software programs, data files, and visualization programs via a Local Area Network (LAN). This collection and integration of software and the migration to a single computer platform results in an approach to LAN/WAN monitoring in either a passive and/or active mode. The architecture permits digital data input from external sensors for analysis, display and correlation with data and displays derived from four major software concept groups. These are: Virus Computer Code Detection; Analysis of Computer Source and Executable Code; Dynamic Monitoring of Data Communication Networks; 3-D Visualization and Animation of Data.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INFORMATION SECURITY ANALYSIS SYSTEM

TECHNICAL FIELD OF THE INVENTION

This invention relates to an information security analysis system for mitigation of Internet security issues and computer source and executable code visualization problems. In addition, the invention relates to an information security analysis system for offensive and defensive Internet analysis and protection of commercial and national infrastructures.

BACKGROUND OF THE INVENTION

Worldwide Internet usage continues to grow at a phenomenal rate. Users include governments, institutions, businesses, and individuals, all of which have connected to the Internet for the purpose of conducting daily activities. Unfortunately, the development and implementation of security measures designed to make Internet connection a secure means of communication have not kept pace with the technological advances in the expansion of network development and interconnectivity. As a result Internet users and networks risk having their information compromised by hackers and malicious users who continue to find ways to exploit and subvert networks and data.

Used appropriately, firewall technologies can help to secure the "front door" of corporate intranets, but these technologies have trouble keeping pace with the applications, services and security that users demand. Although many products have been developed that facilitate network topology discovery, few of these are able to act passively.

Intranet security and monitoring needs are continuing to increase in both government and private industry. This

is substantiated almost daily in trade publications and Internet news groups. More concrete proof of this resides in the increased requirements for security related skills outlined in government requests for proposals. Both government and private industry are spending significant amounts of time and money to address intranet mapping, monitoring, intrusion detection and computer security. This has lead to a prolific amount of organizations, offering to provide intranet computer security services, analysis tools, and associated products.

SUMMARY OF THE INVENTION

The system of the present invention acts passively and provides a methodology for performing a detailed analysis of data observed during a monitoring session.

Without introducing additional traffic on a network, the system of the present invention produces a virtual picture of network usage and network vulnerabilities. By organizing the inputs of multiple collection tools into visual schematics, Security Administrators become proactive in assessing network weaknesses and in identifying optimum locations for implementing security measures. With the information revealed by the system of the present invention, Security Administrators can identify potential traffic bottlenecks, locate the existence of backdoors, reduce bandwidth usage, develop profiles of users, and pinpoint illicit activity.

The software system of the present invention includes four interconnected modules: passive network discovery, network data recording, network data parsing, and network data analysis tools. Network data visualization capabilities are contained within the passive network discovery and network data analysis modules. The software system enables computer code analysis and the 3-D visualization and animation of network traffic and structure. Optional plug-ins further expand and enhance the

software capabilities, thus allowing the software system to remain current regardless of network evolution.

The system of the present invention enables a system administrator to map the network, determine normal and abnormal usage patterns, locate virus attacks, manage network allocation, and display the network.

More technically, the analysis system is a collection, configuration and integration of software programs that reside on multiple interconnected computer platforms. The software, less computer operating systems, are a combination of sensor, analysis, data conversion, and visualization programs. The hardware platforms consist of several different types of interconnected computers, sharing the software programs, data files, and visualization programs via a Local Area Network (LAN). It is this collection and integration of software and the migration to a single computer platform that results in an approach to LAN/WAN monitoring in either a passive and/or active mode. For example, router and firewall software can be monitored in near real time to determine if the code has been functionally changed regardless of security precautions. LAN/WAN data contained in the protocols from the Data Link to Presentation layers in the OSI model are available for analysis with associated displays in two and three-dimensional space.

The architecture also enables digital data input from external sensors for analysis, display and correlation with data and displays derived from four major software groups. These are: Virus Computer Code Detection; Analysis of Computer Source and Executable Code; Dynamic Monitoring of Data Communication Networks; 3-D Visualization and Animation of Data.

The present analysis system templates and displays virus computer code in a graphical functional mode. Current techniques rely on bit stream or real-time monitoring to detect a computer virus in the host computer. The approach of the analysis system of the present invention examines

the functionality of suspect code to determine if a computer virus is present prior to its execution in the host computer. The approach can be viewed as deriving a genetic structure and then determining if the genetic structure is resident, for example, in a computer program, file, or e-mail attachments.

Further, the analysis system of the present invention graphically displays and performs comparisons between like types of computer source and executable code in multi-dimensional space to determine if the code has undergone single or multiple functional alterations. The analysis system enables graphical analysis, code sequencing, and comparison of two or more similar source and/or executable computer programs to determine the degree of functional alteration. This can document, graph, animate, dynamically explore and determine functionality in a single computer source or executable program. The system of the present invention is also capable of sorting source and executable code by language and displaying the results in a graphical functional format. For example, a router's filter table file can be monitored periodically to determine if the file has been functionally changed regardless of current standard security precautions.

The analysis system of the present invention passively discovers the physical and virtual characteristics of digital data communication networks and simultaneously displays different digital communication networks in an interactive manner. Virtual discovery is defined as the ability to determine how the digital data network is being used by its participants and who is connecting to whom at any point in time. This process also determines the configuration changes in a digital data communication network over selectable time intervals. The physical presence of the analysis system of the present invention, in the passive mode, on a LAN/WAN system is undetectable when using conventional techniques, requires no user privileges, consumes no network bandwidth, and does not

interfere with communications on LAN/WAN systems. The analysis system can quickly map SubNets and build complete networks as terminal activity increases. Each active terminal target in the network is mapped and displayed along with appended information. The displayed information shows both physical and virtual relationships, as well as network representations. The analysis system can also be combined with network probes to form remote monitoring, collaboration and discovery of LAN systems. In this scenario, a terminal acts as a master unit with input from the remote probes. In this mode of operation a passive mode of operation may or may not cease depending on whether collaboration is in-band and/or out-of-band.

The analysis system of the present invention dynamically displays, rotates, and animates any data it receives from the three major software groups in three or more dimensions. Simultaneous viewing of different types of digital data in either a physical and/or virtual realms is available.

In accordance with the present invention, the connectivity and functionality for each type of digital data is displayed. The data from each of the three major software groups can be displayed and rotated on any axis on two or more separate but connected visual plains. The process also displays connectivity between different types of data from the three major software groups to include data input from external sensors. The visualization software can append user definable symbols for easier understanding by an operator or analyst. The software interacts with a node via a "mouse click" and dynamically retrieves, decodes and displays information relating to the node that is represented by the three major software groups. In the event that the 3-D nodal diagrams become cluttered, the analyst contracts several nodes into single interconnecting common nodes. This capability provides an uncluttered representation of the original diagram for the

analyst while maintaining functionality of the individual contracted nodes.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the present invention may be had by reference to the following detailed description when taken in conjunction with the accompanying drawings wherein:

FIGURE 1 is a block diagram of an information security analysis system for passive network data discovery, visualization and analysis in accordance with the present invention;

FIGURE 2 is an application flow diagram of the information security analysis system of FIGURE 1;

FIGURE 3 is a block diagram illustrating the architecture for a discovery tool for use with the information security analysis system of FIGURE 1;

FIGURE 4 schematically represents a typical information structure for the discovery tool illustrated in FIGURE 3;

FIGURE 5 is a block diagram of the 3-D visualization module of the information security analysis system of FIGURE 1;

FIGURE 6 is a block diagram of the information security analysis system of the present invention utilized as an intrusion detector;

FIGURE 7 is a block diagram of the information security analysis system of the present invention as an offensive tool for testing for a node attack or information hijacking; and

FIGURE 8 is a typical display illustrating an object-oriented network visualization in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Referring to FIGURE 1, there is illustrated an information security analysis system 10 including a discovery tool 12 for actively or passively monitoring a local access network (LAN) by means of a data channel 14. Functionally, the discovery tool 12 comprises: sensor management, passive network discovery (network viewer, network topology); packet analyzer, knowledge base viewing, and alerting and reporting. In addition, the discovery tool 12 collects traffic and usage data and maps the network connectivity. Data collected by the discovery tool 12 becomes part of a knowledge base 16 stored in memory. Data is organized by major categories as follows: Address, Host, LM-Host, Domain, LM-Domain, SubNet, IP-Address, WWW, MAC-Address, NetwareHost, NetwareNetwork, NetwareStation, Alert, NetwareServer Type, Application, OS, WWW-Browser, WWW-Server, HTTP-Server, NNTP-Server, Protocol, User, POP3-User, FTP-User, SMTP-Sender, SMTP-Receiver, POP3-Password, FTP-Password, Router, and Vendor.

Data in the knowledge base 16 is made available to a data parsing tool 18 that converts the captured network data from the discovery tool 12 to a form useable by downstream programs of the system. Data accessed by the parsing tool 18 is then available to analytical engine 20 for analyzing the data captured by the discovery tool 12 and supports the merging of several data files and the development and comparison of network usage patterns. The analytical engine 20 may be implemented by software from i2 Inc. and marketed under the trademark "Analyst's Notebook". A second analytical engine 20 from the Department of Defense called PROPELLER is also available. The present invention is also capable of utilizing additional analytical engines as such engines become available. The analytical engines 20 are a dynamic set of graphic tools for capturing and displaying a variety of relational data sets in a format referred to as a "link chart". By use of the analytical engine 20, such as "Analyst's Notebook",

data collected can be exploited to characterize and document network characteristics and/or locate possible network intruders. After collecting and organizing data, the analytical engine 20 can be used to make associations between a number of different data charts to determine correlation or differentiation. Relationships between and array of data sources are then available to verify hypothesis, to correlate relationships among multiple data sets and to identify target data within a large data set. Network data needs to be analyzed in order to relate knowledge base data to session data, packet data, and alert data. These relationships assist in determining who has been talking to whom, as well as the content of the traffic for specific protocols (HTP, HTTP, NNTP, POP3, SMTP, TELNET, and IMAP). In the process of analyzing network data, a determination is made as to what IP and/or MAC addresses are common to more than one data set. Characterizing the network in this way, requires taking a periodic snapshot of captured data over a time period. The average of what IP and MAC addresses exists are used to create a link chart representing traffic between each address set. This same process characterizes either a portion of a network or the entire network.

By operation of the analytical engines 20, commonly reused resources may be determined by use of a sampling technique. A time period of interest is identified that will reveal common usage and data is captured during that period. For example, to determine the volume of E-mail traffic between 11:00 a.m. and 1:00 p.m., sampling would occur each day for several weeks until similarities in traffic source and destinations are apparent. After completion of the sampling, the analytical engines 20 can create a chart that inventories all of the IP and/or MAC addresses that have been identified in the sampling.

Several options are available for displaying the analyzed data including a 2-D display 22 and a 3-D display 24. Each of the tools 12 and 18, the analytical engine 20

and the displays 22 and 24 are functionally interconnected to an operator software interface for receiving instructions from an operator of the information security analysis system 10.

The system of FIGURE 1 accepts external sensor data (i.e., biometric information, billing data, SS7, PBX, paging information) in digital formats. When the external data is combined with network discovery and analysis tools there is provided a clear picture of the total communication security process. Thus, the system of the present invention combines physical security needs with electronic communication systems and IS/IT/CIO departments into a complete surveillance package.

In accordance with operator instructions, the system records and plays back selected portions of the stored database for after the-fact analysis and visualization in two or three dimensions. This is adaptable for external sensor and/or intrusion detection data.

In addition, the system of FIGURE 1, may decode FTP, HTTP and TELNET, POP3, SMTP, NNTP, and IMAP sessions in near real time and/or after the fact. The modular architecture of the present invention allows plug-in modules to be added to further enhance and expand protocol decodes to include session reconstruction. This feature permits the analysis system 10 to automatically determine the context of information traveling on an Intranet. This information is then put into nodal diagrams for Network Security personnel to determine what information needs further protection. It also can be used to answer the questions like: are illegal businesses being conducted within the Intranet; what if any harassment is taking place from an employee to another individual; where are employees spending their time on the World Wide web.

In one implementation of the information security analysis system 10, a Pentium-based PC was utilized with a minimum of 166 MHz CPU running the WindowsNT 4.0 operating system. Further, the analysis system 10 included 64

megabyte of real RAM, a 1-gigabyte hard drive and a 17-inch monitor. Improved operation of the information security analysis system 10 is achieved by utilization of a Pentium-II/300 or better with 128 megabyte of real RAM, a 4 gigabyte hard drive and a 21 inch monitor.G13

Referring to FIGURE 2 there is shown a flow diagram of one application of the information security analysis system 10 of FIGURE 1. A passive data discovery engine (discovery tool 12) is utilized to gather data regarding a network and once discovered the data is inserted into the knowledge base 16. Specifically, the discovery tool 12 gathers data to grab small computer source and executable code nodal diagram in two and three dimensional space. Collection of this data enable scaling and displaying large computer code nodal diagrams thereby permitting an analysis the flexibility to view and observe the interconnections within a large body of code for computer equipment that supports digital data communication networks. Gathering computer source and executable code by the discovery tool 12 also enables the system of the present invention to synthetically simulate small computer source and executable code program while viewing related nodal diagram in 3-D space. This enables the determination of where a malicious code might reside within a program, identify memory locations where the data resides after a program has finished execution, and use graphic vectors as templates to find specific types of code modules (that is, viruses, encryption algorithms). In addition the discovery tool 12, collects data on intranets (that is, LAN/WAN) for simultaneous display in two-dimensions the physical and virtual network diagrams. This enables the system analysis to instantaneously display physical equipment net connection of a data communications network. By way of example, by implementing a sum and difference routine, a system analyst is able to determine when new terminals and/or configurations are added or removed from the network to include possible identification of intranet "back-

doors". Collection of this data on internets enables virtual intranet diagrams thereby permitting real time analysis of how the network is being used, who is communicating with whom, determination of potential choke points and vulnerabilities, limited "trace route" reconstruction and types of worldwide web service requested.

In addition, the discovery engine gathers structure information on the network, the method of operation of the network and network users. A typical discovery engine coordinates information from multiple sensors to provide an in-depth picture of network data. In addition, the discovery engine collects data on an operating session of a network in addition to packets of meta data all created along with the knowledge base as "flat files". In addition to gathering and analyzing Ethernet LAN traffic the discovery engine may also be configured to gather and analyze data on other types of network traffic including ATM, WAN protocols, and cellular communications.

The discovery engine (discovery tool 12) generates a knowledge base of data learned about a network and this data is stored in an appropriately named file in a stored data directory of the discovery tool 12. The format of the flat text file from the discovery engine is now processed for further utilization by the information security analysis system 10.

This text knowledge base flat file is processed by the data parsing tool 18 utilizing a keyword search of the knowledge base file to generate data in various categories. For example, data is organized in various categories as follows: unique user identification, host, LM-host, domain, LM-domain, SubNet, IP-address, WWW, MAC-address, NetWare host, NetWare network, NetWare station and various other available categories.

In addition to organizing the knowledge base 16 into various categories, the parsing tool may also create hashed output files.

Following parsing of the knowledge base 16 the analytical engine 20 responds to the data for preparation and converting into vector-based nodal diagrams. Typically the analytical engine 20 creates associations between a number of different charts to determine if such data charts correlate or differentiate. Relationships between an array of data sources is utilized to verify hypothesis, to correlate relationships among multiple data sets, and to identify target data within a large data set. Based on this analysis, the information security analysis system enables the development of resources for management of a network.

The analytical engine 20 analyzes network data to relate knowledge base data to session data, packet data, and alert data as these relationships are utilized to determine who has been talking to whom as well as the content of the traffic for specific protocols.

In the process of analyzing network data received by the discovery tool 12 (discovery engine) a determination must also be made as to what communication exist in more than one data set. Characterizing the data in this way utilizes taking a periodic snapshot of captured data over a time period. Averages are then made of what relationships exist to create a link chart representing traffic between data sets.

Referring to FIGURE 3 there is shown the architecture of a typical discovery tool 12 of FIGURE 1 as illustrated in the application flow diagram of FIGURE 2. One or more sensors are controlled by means of a specialized sensor to provide setup, collection, and transmit control. For the local Ethernet sensor an Ethernet driver sits above the NDIS layer to provide raw packets of network data. Packets of data are queued by a sensor manager 32 and then provided to all the tools in a tool suite 34. An internal packet-processing engine 36 decodes data packets and converts the raw data to information elements that are accessible to all the tools in a tool suite 34. In addition, a script engine

38 filters particularly interesting information and enters knowledge into the knowledge base 16. This database is also accessible by all the tools in the tool suite 34.

In addition to a specialized sensor, the discovery tool 12 also includes control of remote sensor 42. The remote manager 40 queries the remote sensor, for example, a web based monitor and query tool, to be provided to all the tools in the tool suite 34.

As illustrated in FIGURE 3 the discovery tool 12 is organized as a tightly coupled sensor/processor that is based on a suite of inter operable tools. These tools provide visualization, mapping, and analysis of incoming data and processed knowledged. The sensor manager tool 80 provides configuration and control of the sensors within the discovery tool 12 that allows data to be collected (local or remote sensors) without being transmitted to the discovery tool. Various aspects of the sensor manager tool 80 include providing a view of sensors sorted at a top level according to the host, collection of all sensor data within a category, enables transmission of data from sensors to the discovery tool, again by selected category, enables communication from a remote sensor to the discovery tool, adds a new (remote) host and associated sensors to the sensor management tool control.

The network viewer tool 82 provides auto-discovery, auto-layout, and automatic visualization of network nodes and links. Nodes are sources of computer traffic, and include servers, hosts and clients. Links are representations of end to end traffic, and may transfer to higher level network elements (such as routers). The network viewer tool 82 reads packet information and provides a physical picture of one or more logical networks. The logical picture displays nodes and links information and provides a physical picture of one or more logical networks. The logical picture displays node and link information aggregated for multiple packets. Inasmuch as network traffic (nodes and links) exists at many

instances of the OSI network model (data link, etc.), effective visualization occurs by examining the source network at many different layers. In one embodiment of the network viewer tool 82 circles on a graph window represents nodes and lines represent communication links. As the discovery tool 12 automatically discovers more nodes, the count for each network appears on the graph window along with a network label. As the node representation is tree-based, the count is an aggregate of all nodes below the reference node. Information that is relevant to a node from the knowledge base 16 will be displayed in the window of the object viewer tool 84.

The object viewer tool 84 is integrated with the network viewer tool 82, the topology display tool 90, and a query consult tool 94. The object viewer tool 84 actuates the display of information regarding all transitive relations (that are not address-based) that can be made regarding an object. For example, if an IP-address is associated with the user, and a user is associated with a host address, then these will all be a part of the object viewer tool display. However, if the host address is further associated with another IP-address, this transitive association is not displayed because of the confusion that may result in interpreting relations. With nodes being objects and links being relations, the object viewer tool 84 creates a list of objects displayed in a sort by class.

Analysis of data packets and data packet structure is provided by activation of the packet viewer tool 86. This provides the structure of or information within network packets and also helps to discern and understand new, unusual and/or proprietary protocols. When the packet viewer tool 86 is activated, a packet filter (not shown) is initially set to allow all updated packets to be captured. When a user is interested in certain packet types, then the packet viewer tool 86 allows the user to select certain subsets of packets via a packet filter setup dialog. Although the packet viewer tool 86 is useful for protocol

debugging and development, the functionality of this tool also is useful to browse for new packet types.

Turning next to the knowledge browser tool 88, this tool is a visual interface for the knowledge base 16 and provides a tree-based approach to browsing objects in classes within the knowledge base, and in addition provides linkage information for tracing items and information passively discovered on the network by the discovery tool. The knowledge browser tool 88 enables acquisition, organization, and the categorization of network information, of tasks that require both automation for simplicity and customization for user accessibility.

Loosely, a class browsed by the knowledge browser tool 88 is an item in the knowledge base 16 containing categorized information, and can contain subclasses, objects, or both. Examples of classes are IP-ADDR, MAC-ADDR, and SMTP-sender. An object, as considered in the context of the present invention, is a member of a class, and is an item in the knowledge base 16 having network-specific information.

The discovery tool 12 includes the script engine 38 (running as a separate thread) for processing information elements within received protocols to gather intelligence about objects within a network. Standard object types include users, hosts, domains, applications and addresses, however, and ontology specification allows new objects to be added. Using one way or two way bindings to relay information (for example, host and user), associations are made using information elements across multiple protocol/object types. Essentially, in accordance with the function of the present invention, a network becomes a linked graph contained in multi-dimensional space, where relationships are stored as links between vectors within this space.

Next, considering the topology display tool 90, this tool provides a compact, automatically generated view of the elements of a network identified by the discovery tool

12. Figure 8 shows a typical window display upon activation of the topology display tool 90. Based on information contained within the knowledge base 16, the topology display tool 90 shows routers, SubNets, and user nodes. Furthermore, a subset of classes within the knowledge base 16 can be overlaid on top of this view. For example, host names and vulnerabilities can be shown.

The session recorder tool 92 enables packet reassembly, TCP/IP session management, and knowledge discovery. This tool is a mechanism for observing multiple session types which cannot easily be handled at the packet level, for example: HTTP, POP3, SMTP, SNMP, TELNET, NNTP, and IMAP. By reassembling packets and looking for key aspects of information across the reassembled packets, the session recorder tool 92 provides the capability for observing and learning about application level entities on the network.

In operation, the session recorder tool 92 reassembles connection-oriented flows, or sessions. These layer-4 (e.g., TCP) and above sessions consist of multiple packets, to be reassembled and parsed to expose application-level information. Packet and cell reconstruction techniques provide the user with state information (for example, call progress and session monitoring), as well as application layer information (for example, e-mail addresses). Utilizing session search techniques within the session recorder tool 92, combined with alert processing, capabilities (such as seeing when a certain user gets e-mail) can be flexibly constructed. In one implementation of a session recorder tool 92 there is provided viewing of the following sessions: HTTP, POP3, TELNET, FTP, SMTP, NNTP, and IMAP. During operation of the session recorder tool 92 data can be added to the knowledge base 16 as the tool detects, processes and scans sessions for various pieces of information.

The query consult tool 94 provides a text-based interface to the knowledge base 16. By utilization of the

query consult tool 94, a user is able to determine if the knowledge base 16 contains an object (for example, individual IP address) or determine the set of objects belonging to a class of the knowledge base 16 (for example, IP-ADDR). In one implementation of the query consult tool 94, the knowledge base 16 was queried for top level class names, objects belonging to a given class and specific class objects.

In addition to the tool suite 34, the discovery tool 12 includes a knowledge base parsing tool set 96 as shown in Figure 3. Following discovery of the data from the network under analysis, the data is then appropriately formatted for use by the analytical engine 20. The knowledge base parsing tool set 96 functions to take the collected data and put it into the appropriate format for use by the analytical engine 20. Individual tools in the knowledge base parsing tool set 96 are available to parse data from the knowledge base 16 and extract information from saved log files and reassembled session files. The knowledge base parsing tool set 96 comprises eight tools: KB parsing, E-mail extraction, session joining, web extraction, graphics extraction, KB summing, file manipulation, and column splitting.

The network discovery tool generates the knowledge base 16 of data assembled about a network. This data is stored in a flat text file and saved for reuse by the discovery tool 12 for display of a network. The format of the text, however, is not useful for follow on processing. The KB parsing tool, parses the data to be output for display in a columnar file, to be imported to a database, or to the analytical engines 20. In addition, the KB parsing tool is utilized as a key word search to generate data in various categories.

As explained, the session recorder tool 92 is a mechanism for observing multiple session types and generates files containing reassembled session data. The number of files created during a single data collection

may, for example, exceed 10,000. The E-mail extraction tool of a knowledge base tool in the tool set 96 provides for organizing POP3 and SMTP files into summary descriptions. The summary descriptions are then imported to a database or to the analytical engine 20. The E-mail extraction tool contains a key word search mechanism as well as other types of data parsing.

As mentioned, the discovery tool 12 generates a knowledge base of flat file data collected about a network. The web extraction tool of a knowledge base tool set 96 facilitates the parsing and formatting of data from HTML flat files that are then imported to a database or to the analytical engines 20. The web extraction tool contains a tag (similar to a key word) search mechanism as well as other types of data processing algorithms.

The graphics extraction tool of the knowledge base tool set 96 provides for reassembling image files from a recorded format. The display of the session recorder tool 92 provides for the recording of HTTP sessions. These session files contain a header describing the session and the data associated with the session. When a JPG or GIF image is downloaded, the data is reassembled in the session. However, this data is not displayable in the recorded format. The graphic extraction tool converts the reassembled HTTP session file containing JPG and GIF data and creates a new log file containing the names and images.

Data stored in a flat text file by operation of the discovery tool 12 is utilized by the KB summation tool of the knowledge base tool set 96 to create a statistical matrix of the data contained in packet and session logs. For example, the instance of a protocol may be used as the Y access and the source IP address may be used as the X access. After selection of the packet or session log has been made, the KB summation tool screens the appropriate log file and displays available access criteria to create a graph. In the analysis of a typical network, a large number of files will be generated. The file manipulation

tool of the knowledge base tool set 96 provides an interface to reduce the volume of generated files that must be sorted through. It enables files to be deleted or moved based on the file size, type, or contents for purposes of enhancing subsequent processing. Generated files are processed according to a chosen criteria for all files in a group.

Recorded sessions of the discovery tool 12 are occasionally truncated and restored as a new session. These truncated sessions are preferably reassembled before viewing. The session joining tool of the knowledge base tool set 96 connects all truncated sessions into completed transactions.

Also included in the knowledge base tool kit 96 is a split data column tool. This tool is used to remove unwanted columns of data from log files.

Referring to FIGURE 4, there is shown a structuring of the information in the knowledge base 16. The definition and structure of the knowledge is taken into consideration to improve the ability to understand the knowledge prior to processing network information. FIGURE 4 is an organizational chart of categories of information assembled in the knowledge base by the discovery tool 12. The knowledge base is an object-oriented relational entity that is stored as a flat text file and is information collected from packets on the data channel 14.

Referring to FIGURE 5, there is illustrated a block diagram of the 3-D display 24 including a visualization pre-processor 100 receiving raw ASCII data from the analytical engine 20. Also input to the visualization pre-processor 100 through a software link 108 is a visualization setup file 102, a linking information file 104 and a field key file 106. Following processing of the data from the analytical engine 20, the visualization pre-processor 100 transfers the processed data to a 3-D rendering engine 110. The rendering engine 110, a commercial off the shelf software package, formats the

information in accordance with user instructions from an input device 114 and arranges the information for input to a display 112.

Through the use of head mounted display technology and a six degree of freedom tracking system receiving data from the preprocessor 108, a user will experience full viewing immersion within the network identified with the data in the knowledge base 16. This technology provides the user with the further ability to interact and negotiate with the network data display, as opposed to a traditional flat display.

Referring again to FIGURE 1, the 3-D display 24 adds a third dimension to any of the data collected by the discovery tool 12 to view, animate, and analyze complex nodal diagrams in 3-D space. This is required because the raw data file only contains two dimensions. If the data from the analytical engines outputted three or more dimensions, the 3-D display would not be required to add a third dimension. The addition of a third vector permits the simultaneous viewing of large complex diagrams on interconnected planes in accordance with user instructions from the input device 94. The display of FIGURE 5 permits an analyst to rotate the diagram on any axis thereby viewing relationships that otherwise become obscure viewed on two-dimensional planes.

Referring to FIGURE 6 there is shown a representative utilization of the analysis system 10 of the present invention as illustrated in FIGURE 1. The analysis system 10 is operated on a terminal 46 as part of a network including terminals 48, 50 and 52. The network including the terminals 46, 48, 50 and 52 is interconnected through a firewall 54. The firewall 54 interfaces with a network 56 that includes a network analyzer 58. The analyzer 58 analyzes inbound traffic to the terminals and also monitors for "meta data" associated with an intruder inbound to the network. Typically, the analyzer 58 establishes specific meta data associated with an inbound intrusion. As

illustrated in FIGURE 6, the network 56 is coupled to a gateway 60 and a terminal 62 representing a remote intruder to, for example, the terminal 48 as a target.

In the present example it will be assumed that the remote intruder on terminal 60 is attempting to send an E-Mail to the target terminal 48 behind the firewall 54. The analysis system 10 of the present invention running on the terminal 46, monitors through the discovery tool 12 the Ethernet level inbound e-Mail. The analysis system 10 records inbound e-mail traffic as part of the knowledge base 16 such as post office protocol version 3 (POP3)and the simple mail transfer protocol (SMTP). In addition, the analysis system 10 examines meta data associated with inbound E-Mail and further examines SMTP/POP3 packets inbound to the target terminal 48. Identified SMTP/POPS packets inbound for the target terminal 48 are passed to the analytical engine 20 for analysis. As previously explained, the analytical engine 20 imports the meta data passed by the discovery tool 12 for analysis and display.

Referring to FIGURE 7 there is shown a utilization of the analysis system 10 of the present invention in an environment of a multi-node network. As illustrated, the network includes nodes 64, 66 and 68. Interconnected to the node 68 is a terminal 70 running the analysis system 10 as illustrated in FIGURE 1. Also interconnected into the node 68 is a network analyzer 72. Each of the nodes 64, 66 and 68 interconnect to a firewall 74. The firewall 74 in turn is behind an additional firewall 76 that interconnects to a wide area network (not shown).

In this example the analysis system 10 as running on the terminal 70 monitors the level of intranet traffic and records packets of data from each of the terminals of the various nodes. For a terminal under attack, such as terminal 64a, the analysis system establishes a target source packet structure and by means of the analytical engine 20 of the present invention could be modified to shut down a target under attack.

It should be understood that FIGURES 6 and 7 are only two examples of utilization of the analysis system 10. Additional uses of the information security analysis system 10 include offensive and defensive information viewing, context nodal visualization of simultaneous E-mails, FTP and TELNET sessions, graphical playback of filtered nodal traffic, analyzing of computer source and executable code, passively and dynamically discovery of local area network or wide area network physical and virtual connectivity, detection intrusion both internal and external to a local area network or wide area network such as described with reference to FIGURE 6, automatically alert and take corrective action when a network is under attack, FIGURE 7, and detection of computer viruses.

While the invention has been described in connection with a preferred embodiment, it is not intended to limit the scope of the invention to the particular form set forth, but, on the contrary, it is intended to cover alternatives, modifications, equivalents as may be included within the spirit and scope of the invention as defined in the appended claims.

WHAT IS CLAIMED IS:

1. A method for dynamic monitoring of a data communications network, comprising:

gathering information on the physical and virtual characteristics of a data communications network;

generating a knowledge base of the gathered information;

parsing the information in the generated knowledge base to generate data in selected categories in readable format;

analyzing the data in the selected categories to create associations to characterize the data communications network; and

visualizing the analyzed data to determine participant utilization of the data communications network.

2. The method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein generating a knowledge base comprises construction of a nodal network diagram.

3. The method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein generating a knowledge base comprises determining internal and external intrusion attempts.

4. The method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein generating a knowledge base comprises documenting and organizing data in a network functional configuration.

5. A method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein generating a knowledge base comprises generating packets and subpackets of the gathered information.

6. The method of dynamic monitoring of a data communications network as set forth in Claim 1 wherein gathering information comprises gathering information to determine participant use of the communications network at a given point in time.

7. The method for dynamic monitoring of a data communications network as set forth in Claim 6 wherein gathering information further comprises gathering information on configuration changes in the communications network over selectable time intervals.

8. The method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein analyzing the data comprises converting the generated data to vector-based nodal diagrams.

9. The method for dynamic monitoring of a data communications network as set forth in Claim 1 wherein analyzing the data comprises relating knowledge-based data to session data and packet data of the communications network.

10. A method for dynamic monitoring of a data communications networks, comprising:

passively gathering information on the physical and virtual characteristics of nodes and links of the data communications network;

generating a knowledge base of the information passively gathered;

parsing the information in the generated knowledge base to generate data in selected categories in readable format;

analyzing the data in the selected categories to create node and link associations thereby characterizing the data communications network; and

visualizing the analyzed data to determine participant utilization of the data communications network.

11. The method for dynamic monitoring of data communications networks as set forth in Claim 10 wherein visualizing the analyzed data comprises:

storing a visualization setup data file to supply information and data for processing;

storing a network linking data file related to the communications network;

accessing on a selective basis each of the stored files;

transferring the accessed file data to a visualization processor; and

processing the accessed file data and the analyzed data to determine participant utilization of the data communications network.

12. A method for dynamic monitoring of data communications networks as set forth in Claim 10 wherein the analyzed data comprises connectivity and functionality data related to the communications networks and wherein:

visualizing the analyzed data comprises transferring the connectivity and functionality data to the visualization processor;

processing the accessed file data, the connectivity data and the functionality data to generate network data signals for display; and

processing in a 3-D rendering engine the generated network data signals to output network connectivity and functionality display signals.

13. The method for dynamic monitoring of data communications networks as set forth in Claim 12 wherein visualizing the analyzed data further comprises:

generating user input data for control of the 3-D rendering engine;

transferring the user input data to the 3-D rendering engine; and

processing in the 3-D rendering engine the user input data and the generated network data signals to output network connectivity and functionality display signals in accordance with the user input data.

14. The method for dynamic monitoring of a data communications networks as set forth in Claim 10 wherein generating a knowledge base comprises determining internal and external intrusion attempts.

15. The method for dynamic monitoring of a data communications networks as set forth in Claim 10 wherein gathering information further comprises gathering information on configuration changes in the communications networks over selectable time intervals.

16. The method for dynamic monitoring of data communications networks as set forth in Claim 10 wherein visualizing the analyzed data comprises appending user definable symbols to the analyzed data.

17. A method for dynamic monitoring a plurality of a data communications networks, comprising:

gathering information on the physical and virtual characteristics of each of the plurality of data communications networks;

generating a knowledge base of the information gathered;

parsing the information in the generated knowledge base to generate data in selected categories in readable format;

analyzing the data in the selected categories to create associations thereby characterizing each of the data communications networks; and

visualizing the analyzed data to determine the utilization of and the interaction of the data communications networks by participants at any point in time.

18. The method for dynamic monitoring a plurality of data communications networks as set forth in Claim 17 wherein analyzing the data includes remote monitoring, collaboration and discovery of network systems.

19. The method for dynamic monitoring a plurality of data communications networks as set forth in Claim 17 wherein visualizing the analyzed data comprises simultaneously displaying one or more of the communications networks in an interactive configuration.

20. The method for dynamic monitoring a plurality of data communications networks as set forth in Claim 17 wherein analyzing the data comprises mapping subnets to reconfigure networks as terminal activity in a network increases.

21. The method for dynamic monitoring a plurality of data communications networks as set forth in Claim 17 wherein visualizing the analyzed data comprises displaying each active terminal in a network along with nodes and links and appended information.

FIG. 1

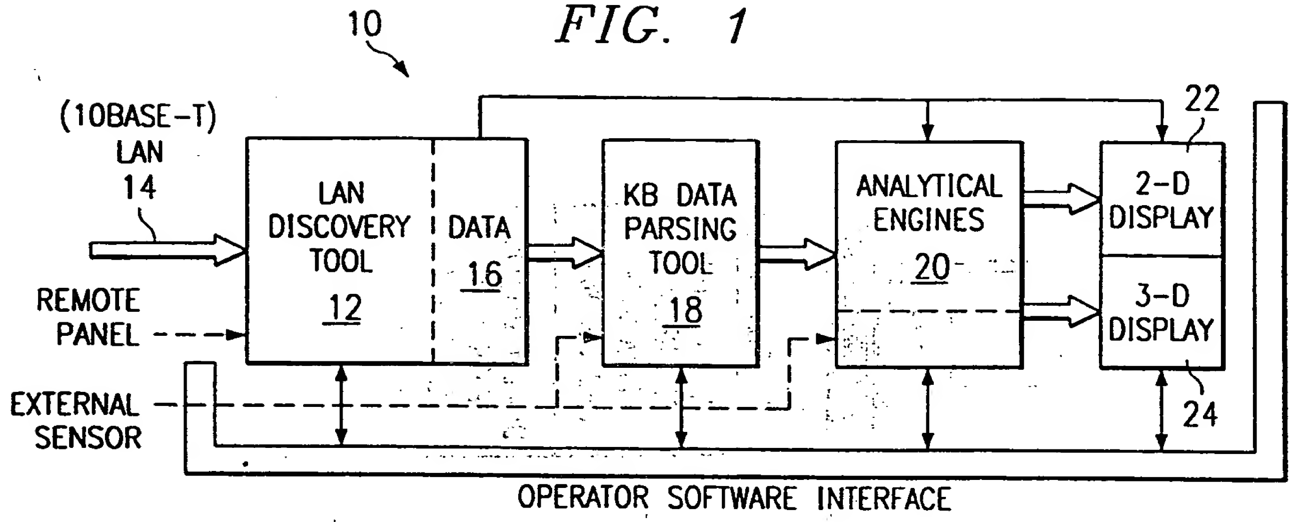

10
(10BASE-T)
LAN
14
LAN
DISCOVERY
TOOL
12
DATA
16
KB DATA
PARSING
TOOL
18
ANALYTICAL
ENGINES
20
2-D
DISPLAY
22
3-D
DISPLAY
24
REMOTE
PANEL
EXTERNAL
SENSOR
OPERATOR SOFTWARE INTERFACE

FIG. 6

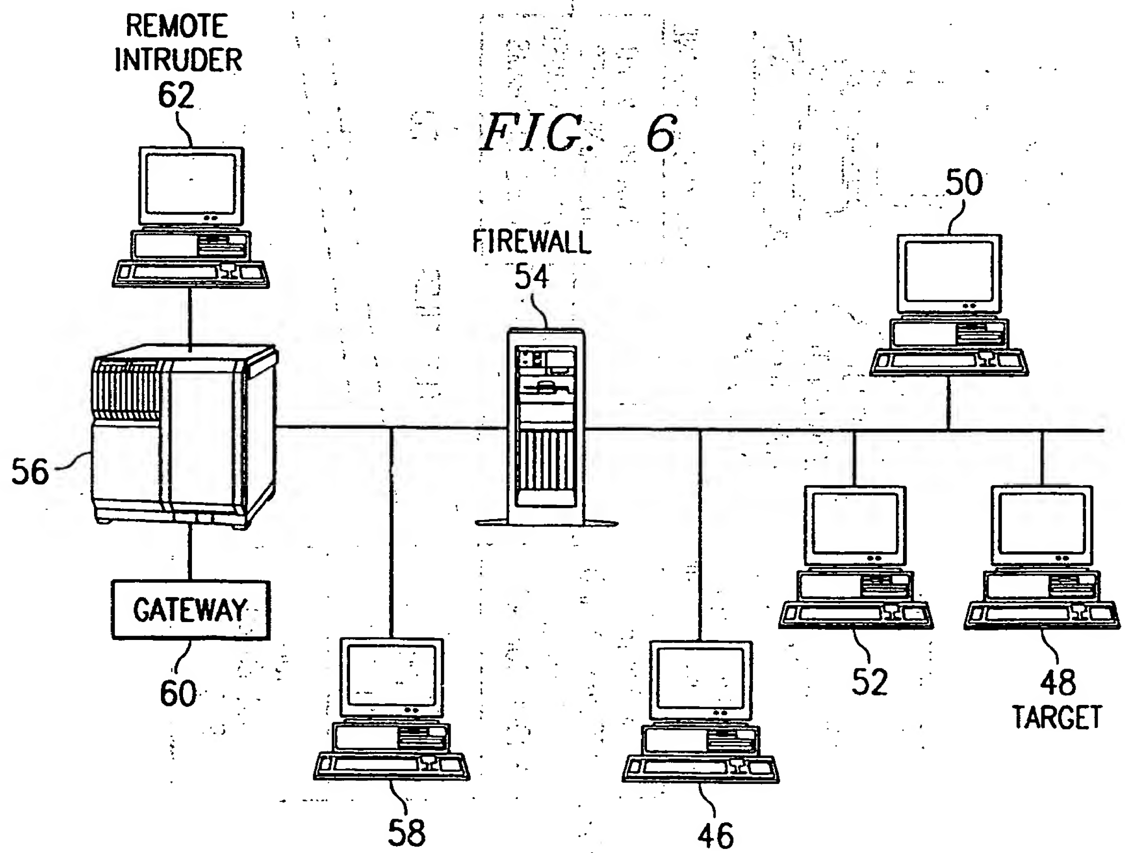

REMOTE
INTRUDER
62
FIREWALL
54
50
56
GATEWAY
60
58
46
52
48
TARGET

FIG. 2

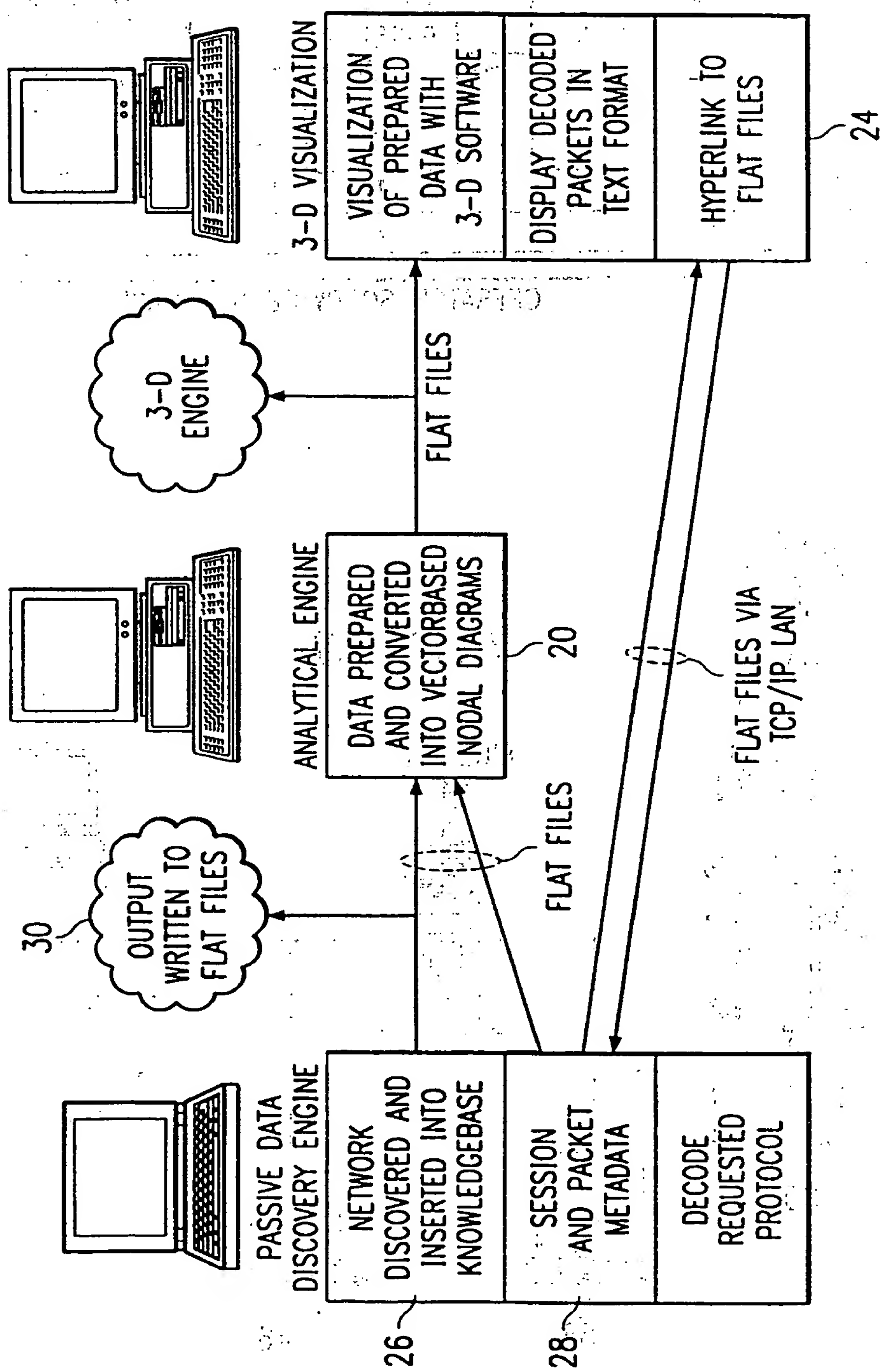
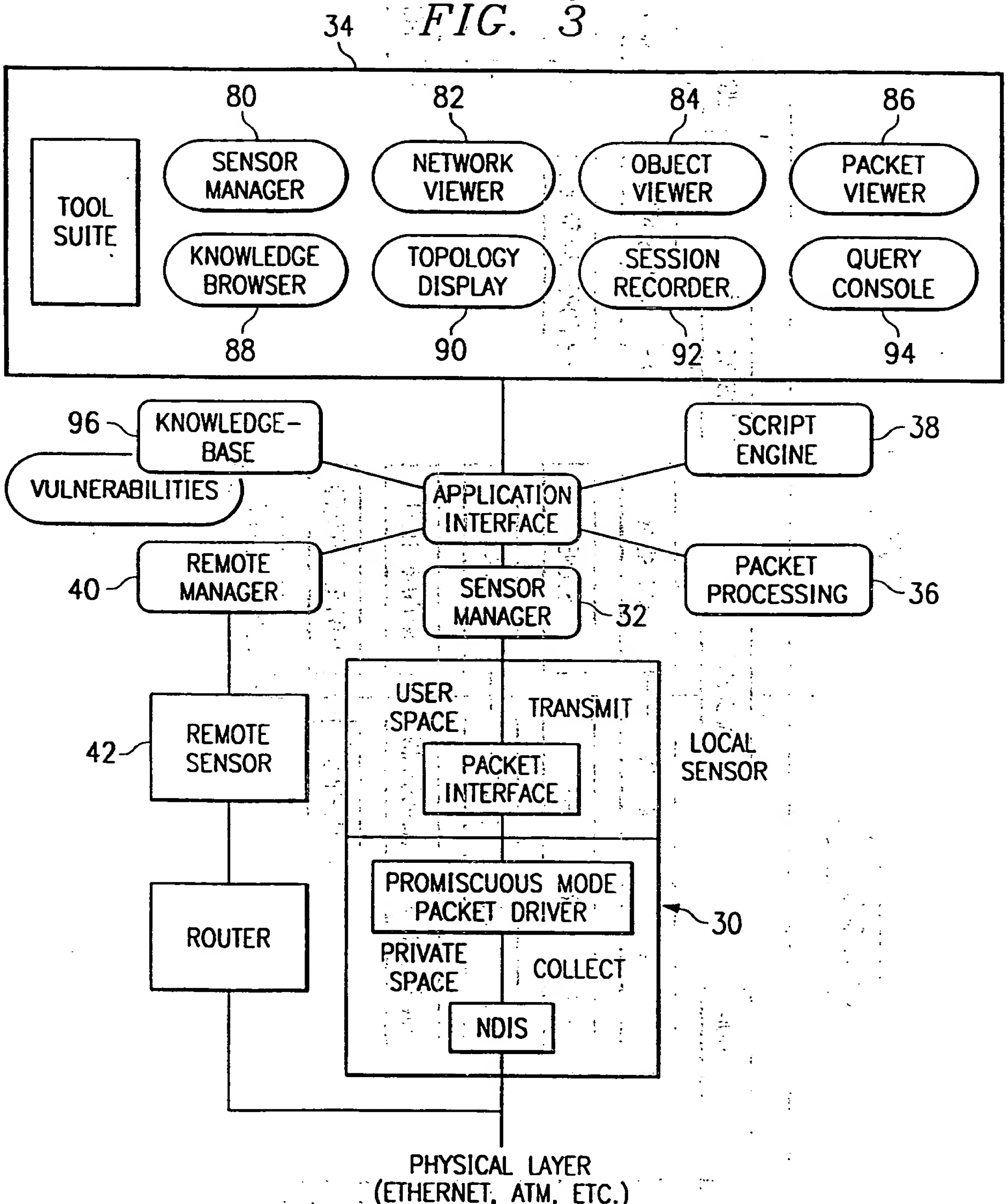

PASSIVE DATA DISCOVERY ENGINE
NETWORK DISCOVERED AND INSERTED INTO KNOWLEDGEBASE
SESSION AND PACKET METADATA
DECODE REQUESTED PROTOCOL
26
28
30
OUTPUT WRITTEN TO FLAT FILES
FLAT FILES
ANALYTICAL ENGINE
DATA PREPARED AND CONVERTED INTO VECTORBASED NODAL DIAGRAMS
20
FLAT FILES VIA TCP/IP LAN
3-D ENGINE
FLAT FILES
3-D VISUALIZATION
VISUALIZATION OF PREPARED DATA WITH 3-D SOFTWARE
DISPLAY DECODED PACKETS IN TEXT FORMAT
HYPERLINK TO FLAT FILES
24

FIG. 3


34
80
82
84
86
TOOL SUITE
SENSOR MANAGER
NETWORK VIEWER
OBJECT VIEWER
PACKET VIEWER
KNOWLEDGE BROWSER
TOPOLOGY DISPLAY
SESSION RECORDER
QUERY CONSOLE
88
90
92
94
96
KNOWLEDGE-BASE
VULNERABILITIES
SCRIPT ENGINE
38
APPLICATION INTERFACE
REMOTE MANAGER
40
SENSOR MANAGER
32
PACKET PROCESSING
36
REMOTE SENSOR
42
USER SPACE
TRANSMIT
PACKET INTERFACE
LOCAL SENSOR
PROMISCUOUS MODE PACKET DRIVER
30
ROUTER
PRIVATE SPACE
COLLECT
NDIS
PHYSICAL LAYER (ETHERNET, ATM, ETC.)

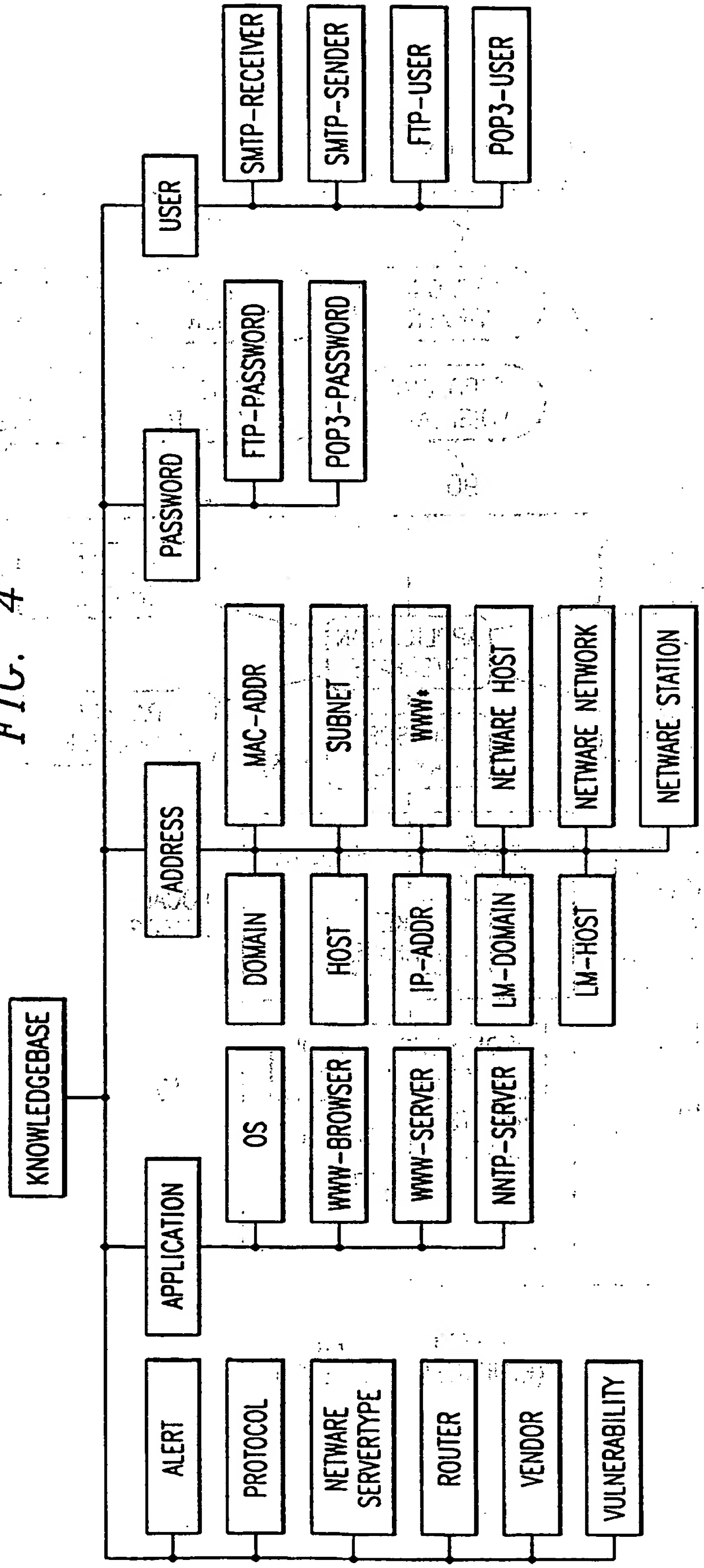

KNOWLEDGEBASE
ALERT
PROTOCOL
NETWARE SERVERTYPE
ROUTER
VENDOR
VULNERABILITY
APPLICATION
OS
WWW-BROWSER
WWW-SERVER
NNTP-SERVER
ADDRESS
DOMAIN
HOST
IP-ADDR
LM-DOMAIN
LM-HOST
MAC-ADDR
SUBNET
WWW*
NETWARE HOST
NETWARE NETWORK
NETWARE STATION
PASSWORD
FTP-PASSWORD
POP3-PASSWORD
USER
SMTP-RECEIVER
SMTP-SENDER
FTP-USER
POP3-USER

FIG. 4

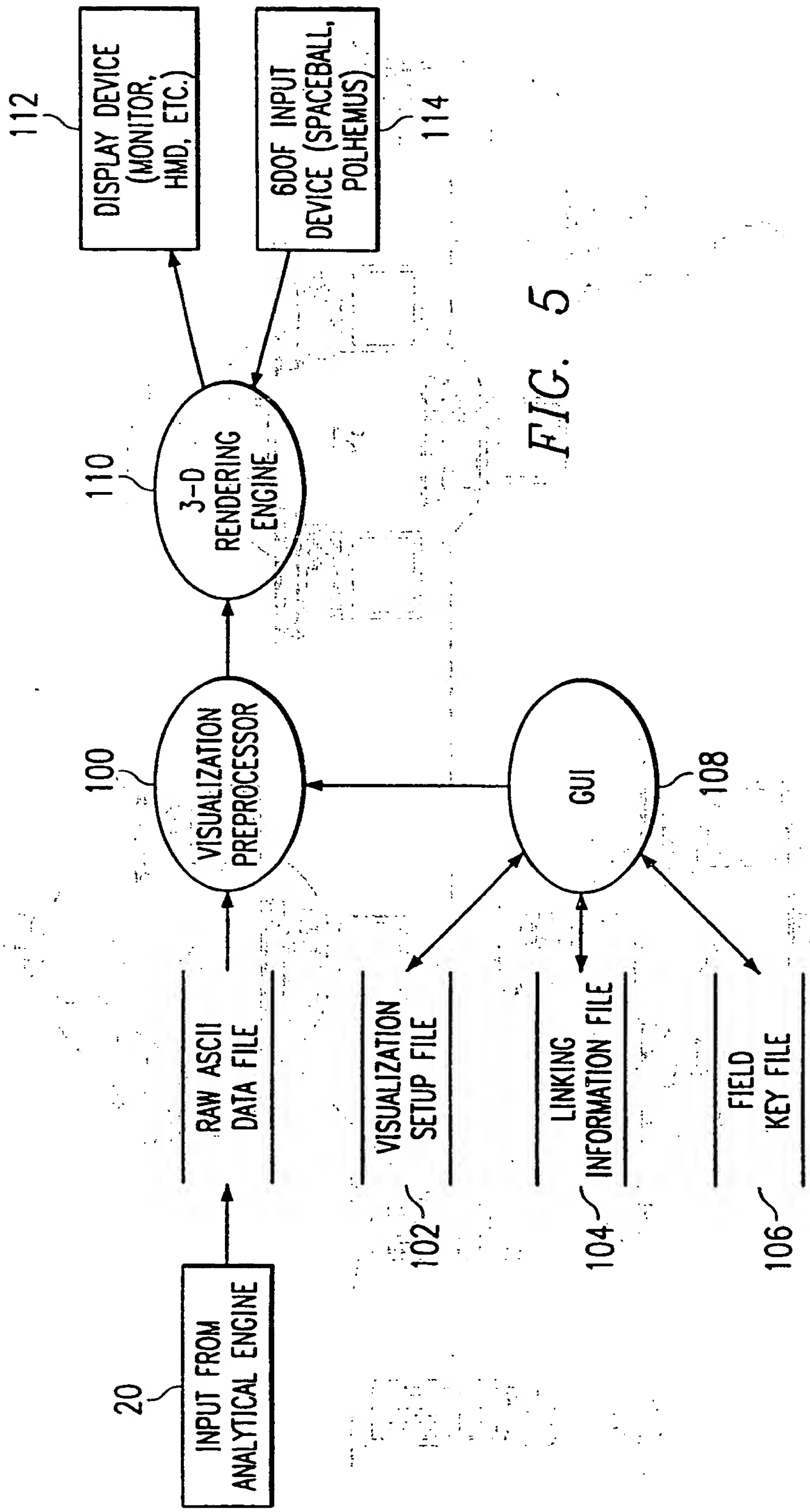

20
INPUT FROM ANALYTICAL ENGINE
RAW ASCII DATA FILE
VISUALIZATION SETUP FILE
102
LINKING INFORMATION FILE
104
FIELD KEY FILE
106
100
VISUALIZATION PREPROCESSOR
GUI
108
110
3-D RENDERING ENGINE
112
DISPLAY DEVICE (MONITOR, HMD, ETC.)
6DOF INPUT DEVICE (SPACEBALL, POLHEMUS)
114

FIG. 5

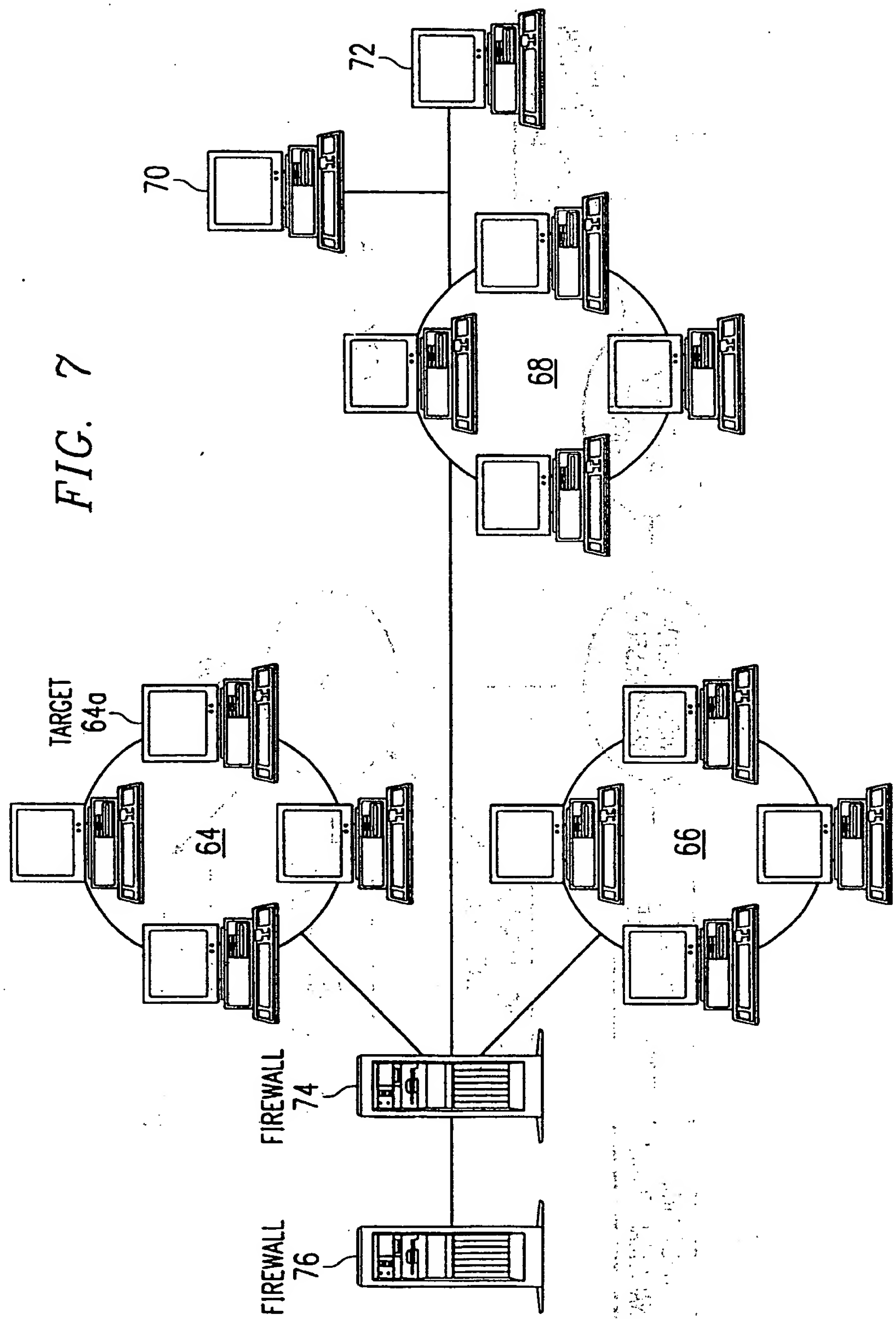


FIG. 7

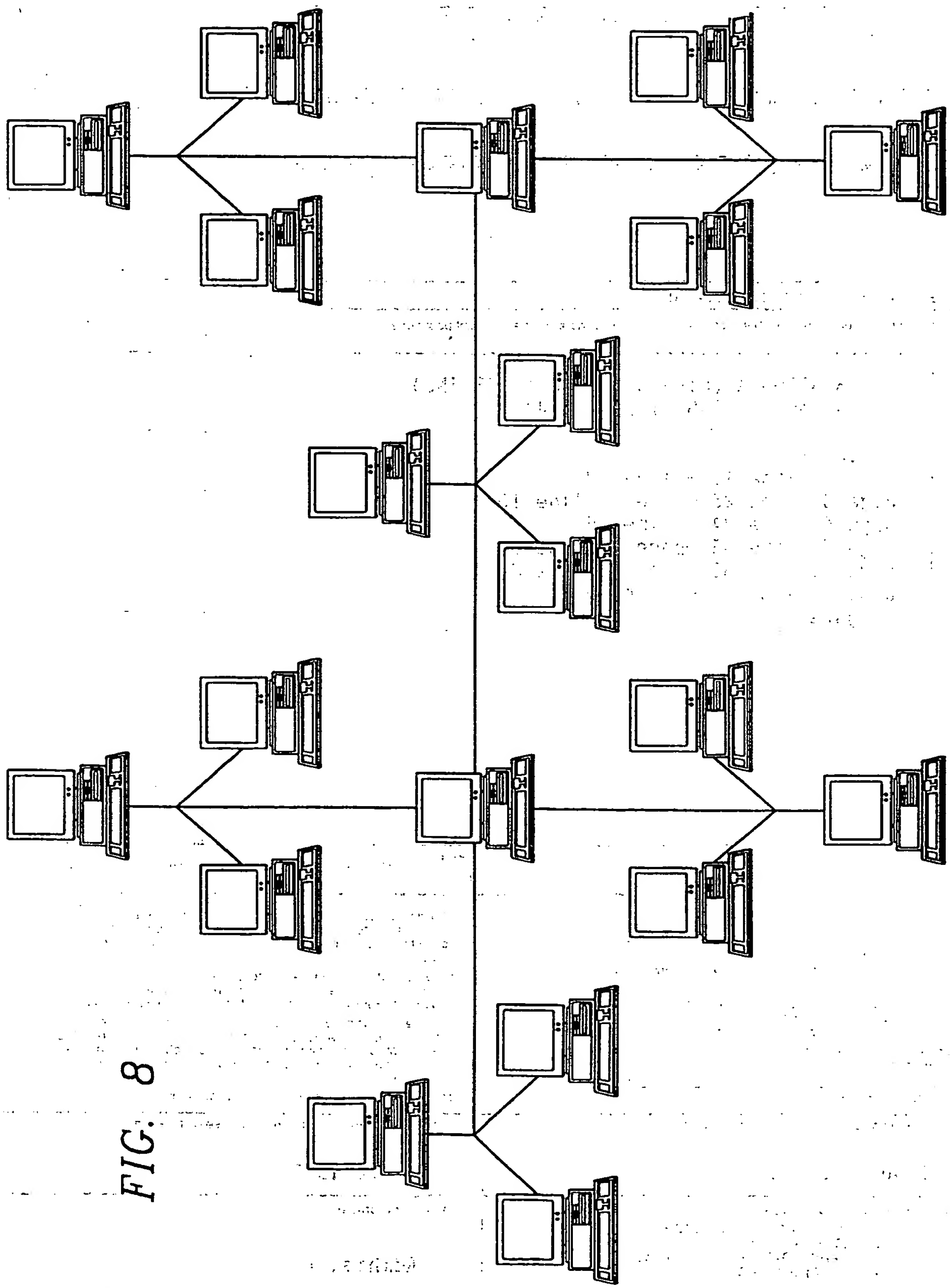


FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/12394

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04L29/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 98 42103 A (FIRSTSENSE SOFTWARE INC) 24 September 1998 (1998-09-24) abstract page 1, line 10 - line 22 page 3, line 26 -page 4, line 10 page 4, line 24 - line 35 page 5, line 35 -page 6, line 9 page 6, line 38 -page 7, line 8 page 7, line 18 - line 26 figure 1 -----	1,3-7,9, 10,14, 15,17,19

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 November 1999	02/12/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Adkhis, F

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte onal Application No
PCT/US 99/12394

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9842103 A	24-09-1998	US 5958010 A	28-09-1999
		AU 6559898 A	12-10-1998